Research on State Prediction Based on Multi-Model Fusion

Yan Lin
School of Information and Science Engineering, University of Jinan, Jinan, 250022, China
Shandong Provincial Key Laboratory of Network based Intelligent Computing, Jinan, 250022, China
Email: jnlinyan@163.com

ZhiQuan Feng*, Bo Yang, Wei Gai and XianHui Song
School of Information and Science Engineering, University of Jinan, Jinan, 250022,China
Shandong Provincial Key Laboratory of Network based Intelligent Computing, Jinan, 250022, China
Email: fzqwww@263.net

***Abstract*—To improve the precision of three-dimensional hand tracking based on the particle filtering, a novel algorithm using Multi-Model fusion to improve the state prediction of particle filtering is put forward. Through analyzing the experimental data of the virtual interactive platform system, combining with the behavior understanding and description of hand, we firstly construct a model for human hand behavior based on cognitive psychology, and then we establishe the state prediction model based on the sigma points during the process of particle filter tracking. We integrate the two models using weighting fusion as the state prediction model during the process of particle filter tracking, which realizes hand three-dimensional tracking based on monocular camera. The results show that the improved algorithm can achieve better prediction precision for hand gestures and improve the tracking precision in the process of hand tracking compared with annealing algorithm. At the same time, since sigma point calculation is simpler, it also saves running time.**

***Index Terms*—particle filter; hands behavior model; sigma points; hand tracking**

I. Introduction

In recent years, target detection and tracking based on video sequences has become the popular research filed of computer vision, pattern recognition, etc. It has important and wide application in target tracking, navigation guidance, condition monitoring, fault diagnosis, parameter estimation and system identification, human arm movement recognition, computer vision, finance areas [1-7] and so on. In order to obtain higher tracking precision during target tracking process, we should establish some models matching with the target motion, and select the appropriate filter algorithm according to the model's characteristic, such as linear and nonlinear etc.

Establishing the model matching with the target motion is one of the two main species of hand tracking, namely, target tracking based on models [8]. Its idea is building a mapping between the information characteristic of models and the information characteristic of images, which needs to acquire the model parameters from the images. This method is applicable to free and complicated Man-computer Interaction task. In this paper, we establish the mapping between the information characteristic of multi-model fusion and the information characteristic of images to acquire the model matching with the target motion.

There exist some limitations to use standard Kalman filtering method for hand three-dimensional tracking, because its posterior probability density is normally appropriated by the first moment and second moment of Gaussian distribution, which is hard to get a good filter effect when the system model is highly nonlinear. Particle filter is out of the constrain that must satisfy Gaussian distribution when solving the nonlinear filtering, can express more extensive distribution than the Gaussian, and has stronger modeling capability for nonlinear property of the variable parameters, so it can effectively solve the problem of nonlinear/non-Gaussian[9-10]. To improve filtering estimation precision, some scholars proposed that we could use particle filter to replace the suboptimal filter in the IMM (Interacting Multiple Model). But with the increase of the state dimension and system models, this simple combination will undoubtedly lead to the rapid expansion of computation. According to the characteristics of particle filter based on random sampling realizing filtered estimate, Driessen H, Mcginnity S et al scholars proposed introducing model information into particle sampling, which realizes joint estimation of states and models in filtering process to reduce the effect of computation with dimension and models increasing [11-12]. Although reduced the computation of dimension and models increasing and improved the speed of sampling, the precision is relatively reduced. To solve the issue and to further improve the tracking precision as the basic goal, this paper proposes a state prediction algorithm based on multi-model fusion and particle filter algorithm. According to the analysis of results compared with annealing algorithm, the new algorithm is verified that it has advantage in hand three-dimensional tracking.

Project number: 60973093, 60873089.
* Corresponding author.

doi:10.4304/jsw.7.2.352-357

II. The State Prediction Model Based on Behavior

A. Three-Dimensional Behavioral Model of Hand

User model is very important in the field of human-computer interaction, it is widely used, and it also has more types. One kind which is called "interface structure model", is starting from the structure of system and discusses the interface's status and decomposition in the system. A typical example is the Seeheim model which is divided the interface into three parts (representation parts, dialogue control, application interface.). The other is "user characteristics model" which understands users from system design. It analyzes the characteristics of different users in order to improve the pertinence and adaptability of the system and to enhance interface personalization and efficiency. A typical example is that users are divided into accident users, strange users, skilled users, expert users according to the user's different understanding of systems, domain of knowledge, experience and skills [13]. From cognitive science, this paper analyzes how the user and computer interactive "interaction model", named behavior model [14].

Three-dimensional behavior model of hand is the analysis of the movement of hands, and describes it in natural language or motion equation. According to BenShneiderman behavioral psychology theory [15], human behavior can generally be divided into seven stages, that is: (1) target defining, (2) intention defining, (3) specific content of action defining, (4) executing, (5) external condition perceiving, (6) external condition explaining, (7) the results of actions assessing. First, observed how people hold objects in real life. Then, based on our observation and BenShneiderman behavior psychology theory, we divide the process of people holding objects into six steps: (1) target defining (e.g. hand bring the object), (2) hand is moving to the target, (3) grasping the object, (4) hand and the object moving to the destination as a whole, (5) putting the object on the destination and losing, (6) return to the natural state.

B. Cognitive Experimental Platform

Based on the process of hand holding objects, we establish a basic virtual hand interactive platform, e.g. figure 1. The platform is human-computer interaction which uses the Data Glove (5DT Data Glove 14 the Ultra) and position tracking devices in the virtual space of OpenGL rendering. Through the sensor of the data glove and position tracker, the computer can obtain the hand information on time, it firstly reconstruct the three-dimensional hand in the virtual hand interaction platform, and then control the hand doing the gesture corresponding the real hand to complete the interactive action of hand holding object. We record the hand data of the whole process for analyzing.

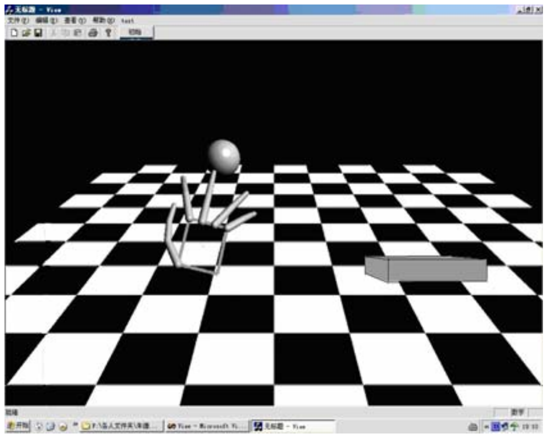

Figure 1. Virtual hand interaction platform

C. State Prediction of Hand Behavior Model

According to the skeletal structure of the hand, we describe the hand gesture with 33 DOF (Degree of Freedom), including nine global freedoms (the center position and direction of the palm) and 24 partial freedoms [16] (mainly refer to motion parameters of each joint: the position relationship of the finger and the palm, the position relationship of each finger.). Operators do the experiment in accordance with the basic process of hand holding object in the virtual hand interaction platform. After being trained skilled, operators do *n* times experiment and record the hand data, and take the mean of the *n* times experimental data. Figure 2 is the mean value of *n* times experimental data of 24 partial freedoms in the whole process of hand holding object. In fig.2, the horizontal axis stands for the frames of the whole process of holding object, and the vertical axis stands for the values of 24 partial freedom, and 1, 2……24 are 24 partial freedom degrees of hand. Observed from the figure, in the whole process of hand holding object, each joint data of hand changes only in the process of grasping and loosing of the hand, and in the other part of process it is unchanged. The two changing data approximate parabolic curve and the other data approximate linear. Therefore, according to the characteristic of the data, we use quadratic polynomial and the principle of least square method to do piecewise fitting. The process of hand holding object can be expressed by mathematical expression,

$$y_i^l = a_i^l x^2 + b_i^l x + c_i^l \tag{1}$$

a_i^l , b_i^l , c_i^l are the coefficient of the i^{th} partial freedom of hand in the l stage of hand holding object. a_i^l, b_i^l, c_i^l are fitted by quadratic polynomial. *x* refers to the number of frame, y_i^l is parameter values of the i th partial freedom of hand in the l stage of hand holding object, $0 < i \le 24$. Formula (1) will be a model of hand state prediction.

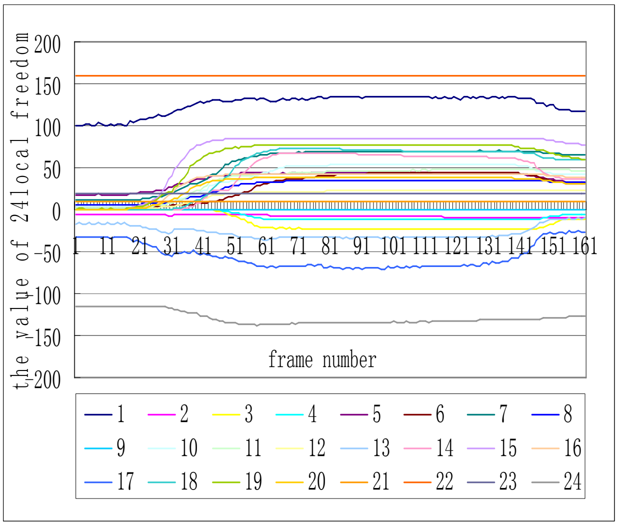

Figure 2. 24 partial freedom value changes in the process of hand holding object

III. The State Prediction Model Based on Sigma Point

A. Sigma point definition and acquisition

X is a known random vector, its mean is $\bar{X}$, covariance matrix is P_x, and $Y = f(X)$. It is assumed that there are M weighted vector sets $S = \{(X_i, \omega_i) \mid \sum_{i=0}^{M} \omega_i = 1, i = 1,2......M\}$ (i=0, $\bar{X}$ refers to itself)around (including $\bar{X}$), and S has the same mean and the same covariance matrix with X:

$$Y_i = f(X_i), \tag{2}$$

$$\bar{Y} = \sum_{i=0}^{M} \omega_i Y_i, \tag{3}$$

$$P_Y = \sum_{i=0}^{M} \omega_i (Y_i - \bar{Y})(Y_i - \bar{Y})^T, \tag{4}$$

Then every vector of the S is called a sigma point of X.

In this experiment, we consider the movement of hand to approximately obey Gaussian distribution, so we choose symmetric strategy to sampling. Only the mean and covariance being taken into account, we use L sigma points to approximate X, $L=2n+1$, where n is the dimension of X, then X can be denoted as:

$$X = \bar{X} + CZ \tag{5}$$

Where, Z is a random variable of standard normal distribution, C is RMS (Root Mean Square) matrix of P_x, namely, $P_x = CC^T$, C is easily to find using Cholesky method. Based on the definition of sigma points, we can get the formula for sigma point [17].

$$\begin{cases} X_0 = \bar{X}, w_0 = 1 - \dfrac{n}{3}, \\ X_i = \bar{X} + \sqrt{3}(\sqrt{P_X})_i, w_i = \dfrac{1}{6}, \\ X_{i+n} = \bar{X} + \sqrt{3}(\sqrt{P_X})_i, w_i = \dfrac{1}{6}, \end{cases} \tag{6}$$

B. State Prediction Based on Sigma Point

From the definition of sigma point, the mean and the covariance of X can accurately achieve the second order in the process of finding sigma points, the result of experiment can get a relatively higher filtering. Thus, in order to improve precision, in this paper, we establish state prediction model based on sigma points by using the data of tracking process. During the tracking process, we take n frames data of hand three-dimensional tracking before the current frame, and get their mean $\bar{X}$ and covariance matrix P_x. We can get $2n+1$ sigma points of hand three-dimensional data by the (6). In this paper, we make an assumption that the data of hand tracking obeys Gaussian distribution, the position and the weight of sigma points have nothing to do with the non-linear transformation[17], thus $Y = X$, we can obtain:

$$\begin{cases} Y_i^k = \bar{X}_i^k, k = 0 \\ Y_i^k = \bar{X}_i^k + \sqrt{3}(\sqrt{P_x})_i^k, k = 1,2,3......2n \end{cases} \tag{7}$$

Where, i is the i^{th} partial freedom of hand, $0 < i \le 24$. Formula (7) is the state prediction equation based on sigma point.

IV. State Prediction Algorithm Based on Multi-Model Fusion

Since the three-dimensional movement of hand is a typical nonlinear and non-gaussian movement, this paper uses particle filter to express the posteriori probability distribution based on the observed quantity and the controlled variable, and approximates probability density function by finding a set of random samples in the state space. In this paper, we integrate the two models using weighting fusion, which is applied to the state prediction process of particle filter to improve the precision of state prediction, so as to improve the precision of hand tracking.

A. The analysis of the main data flow in the algorithm

Since the data in the first process of hand holding object approximate linear, the movement of hand is assumed to be uniform motion, and according to the previous frame state, we can get the current frame prediction state based on continuous deformation. So we get the first n frames data based on the status of previous frames data, when initialization is success. As the first n frames data are all obtained, we construct the state prediction model based on the sigma points during the process of particle filter tracking. Then we integrate the state model based on behavior and the state model based on sigma points into a model, which is the state prediction model during the process of particle filter tracking. Figure 3 is the analysis of the main data flow.

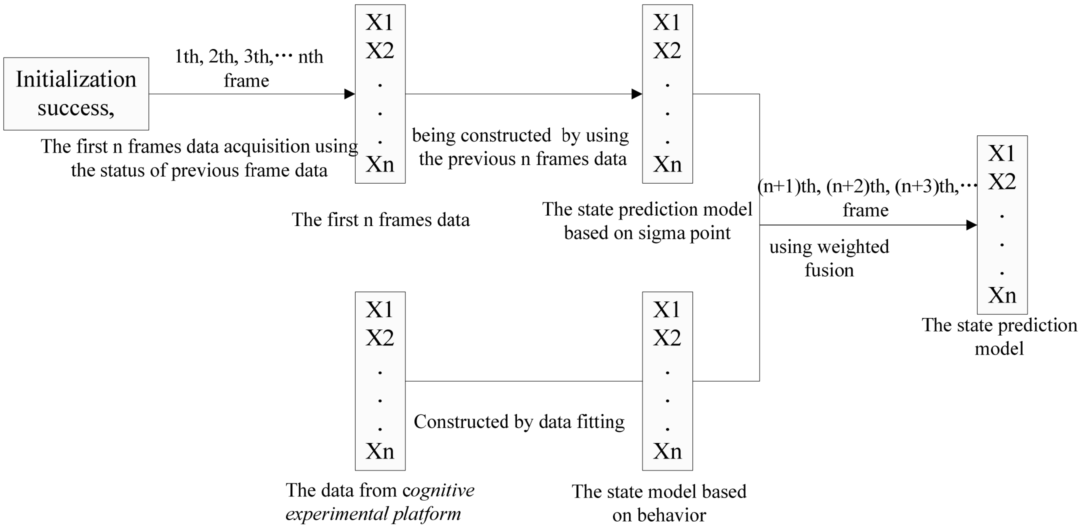

Figure 3. The analysis of the main data flow.

B. *Algorithm Description*

1) Steps:

a) Initialization: This paper uses a dynamic initialization method of sport hand [16].

b) State Prediction: The state prediction equation y_i^l of hand behavior model and Y_i^k based on sigma point are weighted fusion, namely:

$$Yy_i = a_i^l y_i^l + b_i^k Y_i^k \quad (8)$$

Where, Yy_i is the state prediction value of the i^{th} partial freedom of hand, a_i^l , b_i^k is empirical value.

c) Particle Sampling: This paper uses Gaussian sampling, the number of sample particles are *2n+1*, and *2n+1* is the number of sigma points,

$$\begin{bmatrix} YY_i^1 \\ YY_i^2 \\ \vdots \\ YY_i^{2n+1} \end{bmatrix} = \begin{bmatrix} Yy_i^0 \\ Yy_i^1 \\ \vdots \\ Yy_i^2 \end{bmatrix} + \begin{bmatrix} u_G \\ u_G \\ \vdots \\ u_G \end{bmatrix} \quad (9)$$

Where, u_G is the value which is generated by Gaussian sampling. $YY_i^1, YY_i^2 ... YY_i^{2n+1}$ is the particle value of the i^{th} partial freedom of hand after particle filtering, $Yy_i^0, Yy_i^1 ... Yy_i^{2n+1}$ is the state prediction equation, and

$$Yy_i^j = a_i^l y_i^l + b_i^k Y_i^j, j = 0,1,2...2n+1 \quad (10)$$

d) Particle Weight: Using the particle weight in this paper, we calculate the similarity degree between the particles produced during the state prediction process and the gesture images which are obtained during the movement process of real hand. The similarity degree is measured by the distance of hand particles between the projection *M* of the image plane and the frame image *N* of the current frame, and it can be expressed as follows:

$$\omega_i = \frac{1}{Hausdorff(M, N)} \quad (11)$$

And normalize the weights:

$$\omega_k^i = \frac{\omega_k^i}{\sum_{j=1}^{2n+1} \omega_k^j} \quad (12)$$

e) State Estimation: The final estimated state of the particle is integrated with every sampling particle and their weight information. Estimation formula is defined as:

$$YY_k^i = \sum_{j=1}^{2n+1} YY_k^j \omega_k^j, i = 1,2,3......33 \quad (13)$$

V. EXPERIMENTAL RESULTS AND ANALYSIS

The experiment in this paper is done on the PC, whose CPU is Intel(R) Core(TM) 2, frequency is 2. 66GHz, and memory is 4GB.

A. *Experiment 1*

In order to verify the state prediction value being to combine information from multiple models, we track the parameters of forefinger joint during the second process of hand holding object. In this paper, we do eight experiments using our algorithm, record tracking data of forefinger joint freedom per frame, and take the mean of eight experimental data compared with the average data of cognitive experimental platform. Figure 3 is the data comparison of forefinger joint freedom. It indicates that there is a great deal of difference between the two curves. This is because state prediction value is not a single model data but comprehensive results of each model, which is more consistent with the state changes in the motion process of human hand. Figure 4 is the experimental simulation.

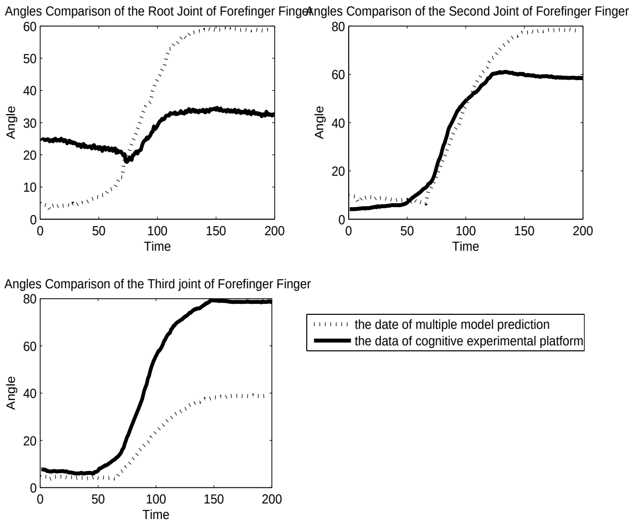

Figure 4. The data comparison of forefinger tracking and cognitive experiment

B. *Experiment 2*

In order to verify the effectiveness and superiority of improved particle filtering, our algorithm is compared with annealing algorithm from the average tracking precision and the average running time. The evaluation of tracking precision is measured by Hausdorff distance, the formula is defined as:

$$p = \exp(-kh)$$

(14)

Where, k=0.01, h is Hausdorff distance, which can be obtained by (11).

We do ten times independent experiments respectively using our algorithm and annealing algorithm. In our algorithm, we take five dimensional state data, and get eleven sigma points, so the number of particles sampling is eleven. Figure 5 is the comparison of tracking precision. And it shows that improved particle filter can achieve good precision.

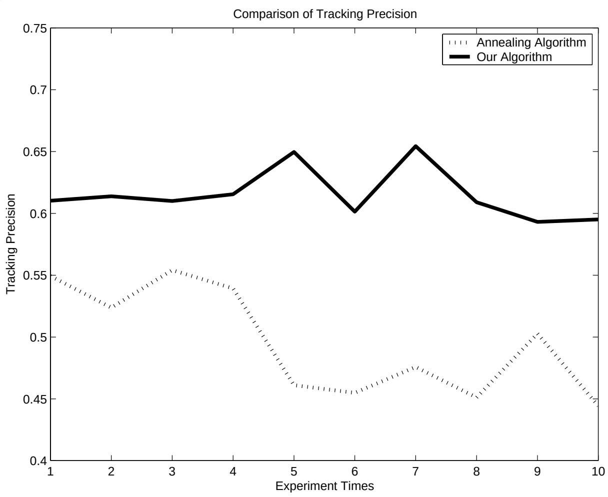

Figure 5. The comparison of tracking precision

Figure 6 is the comparison of average running time per frame. It indicates that the times consumed of our algorithm per frame are less than the times consumed of the annealing algorithm, that is because the calculating process of sigma points is relatively simple. Thus, our algorithm saves the times in the hand tracking process, it has good real-time tracking.

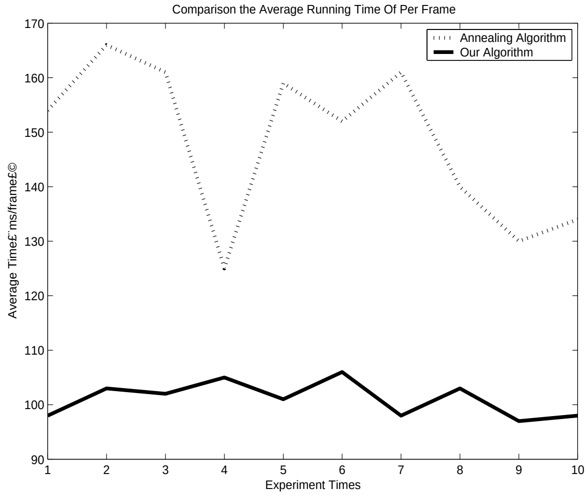

Figure 6. The comparison of average running time per frame

Figure 7 is the experimental simulation. It shows that the hand in the three-dimension is similar to the hand which is obtained by the camera.

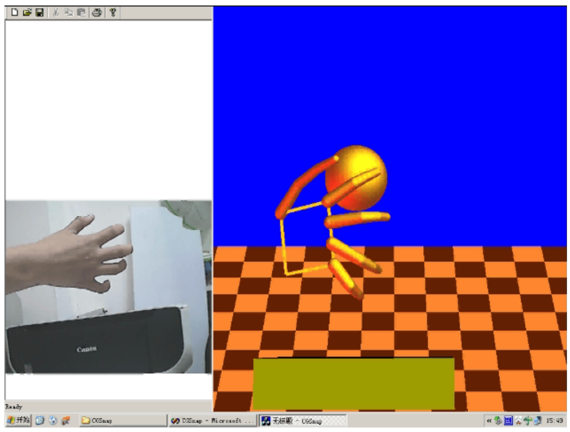

Figure 7. Simulation in the experiment

VI. CONCLUSION

In this paper, we combine the characteristics of hand with hand behavioral model from the perspective of cognitive psychology. In the process of hand tracking, we establish the state prediction model based on sigma point. The particle filter algorithm is improved based on the fusion of the two models fusion, which can modify the precision of hand tracking by using the data during the tracking process to correct the state prediction value. At the same time, since the calculating process of sigma point is relatively simple, our algorithm saves the times in the hand tracking process, and it also has good real-time tracking. However, it is undeniable that there are still some problems and limitations. One is that cognitive behavioral model must be established under certain conditions. As this article uses the data glove and position tracker equipments to get the data in the basic process of hand holding object. And if there are not these devices, what should we do. It requires us to further study. Another is that the application scope of hand behavioral model is limited, and it needs to be further expanded.

ACKNOWLEDGMENT

This paper is supported by National Natural Science Foundation of China (No.60973093, No. 60873089), Natural Science Foundation for Distinguished Youth Scholar of Shandong Province (No.JQ200820), Key Project of Natural Science Foundation of Shandong Province (2006G03), Science and Technology Plan of Shandong Province Education Department (J07YJ18).

REFERENCES

[1] Godsill S.J, Vermaak J, Ng W and Li J.F. “Models and Algorithms for Tracking of Maneuvering Objects Using Variable Rate Particle Filters,” Proceedings of The IEEE, Vol.95,No.5, 2007.pp.925-952.

[2] Carmi A and Y. Oshman, "Adaptive Particle Filtering for Spacecraft Attitude Estimation from Vector Observations," Journal of Guidance, Control and Dynamics, Vol.32, No. 1, January - February 2009. pp. 232 - 241 .

[3] Tafazoli S and Sun XH. “Hybrid System State Tracking and Fault Detection Using Particle Filters,” IEEE Trans. On Control Systems Technology, Vol.14, No.6, Nov.2006. pp.1078.

[4] Saha, Nilanjan and Roy D. “A Girsanov Particle Filter in Nonlinear Engineering Dynamics,” Physics Letters A, Vol.373, No.6,2009. pp.627-635..

[5] Chang WY, Chen CS and Jian YD. “Visual Tracking in High-Dimensional State Space by Appearance- Guided Particle Filtering,” IEEE Trans Image Processing, Vol.17, No.7, Jul.2008. pp.1154-1167.

[6] Chang C and Ansari R.Kernel. “Particle Filter for Visual Tracking,” IEEE Signal Processing Letters, Vol.12, No.3, 2005. pp.242.

[7] Aiube FAL, Baidya TKN and Tito EAH. “Analysis of Commodity Prices With the Particle filter,” Energy Economics, Vol.30, Mar.2008.pp.597-605.

[8] Lee J and Kunii T. “Model-based analysis of hand posture,” IEEE Computer Graphics and Applications archive.Vol.15, No.5, Sep.1995..

[9] Musicki D and Suvorova S. “Tracking in Clutter Using IMM-IP-DA-based Algorithms,” IEEE Transactions on Aerospace and Electronic Systems,Vol.44, No.1, Jan.2008. pp.111-126.

[10] Mallick M and La Scala B F. “IMM Estimator for Ground Target Tracking with Variable Measurement Sampling Intervals,”//9th International Conference on Information fusion. 2006. pp.1-8.

[11] Driessen H and Boers Y. “Efficient Particle Filter for Jump Markov Nonlinear Systems,” IEE Proceedings Radar Sonar Navigation, Vol.152,No. 5, Oct.2005.pp.323-326.

[12] Mcginnity S and Irwin G W. “Multiple Model Bootstrap Filter for Maneuvering Target Tracking,” IEEE Transactions on Aero-space and Electronic Systems, Vol.36, No.3, Jul.2000. pp.1006 - 1012.

[13] DONG Shi-hai. “Progress and Challenge of Human-Computer Interaction,” Journal of Computer-Aided Design & Computer Graphics, Vol.16, No.1, Jan.2004. pp.1-11. (references)

[14] WANG Xing-peng. “Summarization of Cognitive Model,” Computer Engineering and Design, Vol.28, No.16, 2007. pp.4009-4011.

[15] LIANG Ling-jian. “Modern Cognitive Psychology,” Shang-hai: Shanghai Education Press,2006.(in Chinese)

[16] LI Yi, FENG Zhiquan, YANG Bo and SONG Xiaona. “Human Hand 3D Tracking Based on Particle Filter,” Journal of Southeast University (Natural Science), vol.40, Sup(I), Sept.2010. pp.172-177. (in Chinese)

[17] FENG Zhiquan, YANG Bo, LI Yi, ZHEN Yangwei and ZHANG Shaobai. “Research on Human Hand Tracking Aiming at Improving Its Accurateness,” Journal of Computer Research and Development, Vol.45, No.7, 2008. pp.1239-1248. (references)

Yan Lin, born in 1989, enrolled postgraduate. Her main research interests in moving hand tracking and hunman-computer interaction.

ZhiQuan Feng, born in 1964. Ph.D, Associate professor, and maste supervisor. His main research interests include image processing, moving human hand tracking, and human-computer interaction.

Bo Yang, born in 1965. Ph.D., Professor And Ph.D. supervisor. His main research interests include eimage processing and Intelligent network information processing.

Wei Gai, born in 1987, enrolled postgraduate. Her main research interests on human-computer interaction and moving human tracking.

XianHui Song, born in 1986, enrolled postgraduate. Her main research interests on human-computer interaction and moving human tracking.